Research on Improved PPLCNet Classification Network Based on CBAM Attention Mode

Peng Wang*

College of Electronic Information and Automation, Tianjin University of Science and Technology, 300222, China

Shengfeng Wang

College of Electronic Information and Automation, Tianjin University of Science and Technology, 300222, China

Qikun Wang

College of Electronic Information and Automation, Tianjin University of Science and Technology, 300222, China

Yuting Zhou

College of Electronic Information and Automation, Tianjin University of Science and Technology, 300222, China

*E-mail: *2324365941@qq.com, autowangpeng@tust.edu.cn, globeen@163.com,*

2994383067@qq.com

www.tust.edu.cn

(Tianjin University of Science and Technology, Tianjin, China)

Abstract

This paper studies pedestrian attribute recognition based on the pplcnet network because it is of great significance in the field of traffic security. Firstly, the research status of pedestrian attribute recognition and common deep learning models are introduced. Secondly, considering that convolutional block attention module contains both spatial attention module and channel attention module, we add this attention model to pplcnet to improve performance. Finally, this paper verifies the model through the pa100k dataset and obtains good results.

Keywords: Neural networks, Pedestrian attributes,CBAM, Deep learning

1. Introduction

In the current context of rapid technological advancement, the significance of attribute recognition technology is increasingly prominent [1]. It plays a crucial role in key sectors such as healthcare, public security, and intelligent furniture design. This has drawn extensive attention from academia and industry alike.With the deep integration of deep learning algorithms , vast amounts of data can be effectively processed. Computer systems can utilize these data via continuous self-learning.Notably, the person attribute recognition technology based on deep learning stands out. Through convolutional and pooling neural networks, it can precisely extract feature data from pedestrian images.Via complex classification procedures, it can acquire multi-dimensional pedestrian attribute information. The extraction of such attribute information enriches our understanding of individual characteristics. It also lays a practical foundation for applications in smart city construction and other crucial areas like military security. This promotes technological innovation and development transformation in related fields [2].

The rest of this article is organized as follows. The second section of the article introduces the deep separable convolutional network, pplcnet and data sets. In the third chapter, this paper compares the performance of the network by focusing on the location of the network. Finally, in Chapter 4, we summarize the full text.

2. Neural Networks and the Principle of Attention

In the field of computer vision, deep learning has become a basic tool. In this study, a large number of deep learning concepts will be used, and deep learning neural networks will be used to identify pedestrian attributes. Finally, results will be obtained in experiments.

2.1. Depthwise separable convolution

Depthwise Separable Convolution (DepthSepConv) is an efficient convolution operation method, which plays an important role in the field of deep learning. Especially in the architectures of deep convolutional neural networks.

Structurally, it consists of two crucial steps: depthwise convolution and pointwise convolution. In the depthwise convolution stage, for the input data. Such as a feature map with the shape of $H \times W \times C$ (where H represents the height, W represents the width, and C represents the number of channels). Each convolution kernel processes the data of a single channel independently.

In the pointwise convolution stage, a 1×1 convolution operation is used to fuse the features of each channel obtained in the depthwise convolution stage. This operation can change the number of output channels, enabling the network to adjust the dimension of the feature representation .

In terms of performance advantages, DepthSepConv performs excellently in terms of the amount of computation and the number of parameters. Compared with traditional convolution, its amount of computation and the number of parameters are significantly reduced. Taking traditional convolution as an example, the amount of computation is approximately $H \times W \times C \times K \times K \times N$, and the number of parameters is approximately $C \times K \times K \times N$. For DepthSepConv, the amount of computation is approximately $H \times W \times C \times (K \times K + N)$, and the number of parameters is approximately $C \times (K \times K + N)$.

2.2. Practical plain - based lightweight convolutional neural network

PPLCNet (Practical Plain-based Lightweight Convolutional Neural Network) is a lightweight convolutional neural network model that has drawn significant attention in the field of deep learning. It is designed to strike a balance between model accuracy and computational efficiency, while the storage requirements are minimized. This makes it highly suitable to be deployed in resource-constrained devices and applications with strict real-time constraints.

The network architecture of PPLCNet is characterized by several key features. A specialized depthwise convolution structure is employed. Compared with traditional convolutional layers, the number of parameters can be reduced by performing convolutions on each channel independently in the depthwise convolution and then integrating the channel information via pointwise convolutions (1×1 convolutions). In this way, the fine-grained local features, such as the texture and edge details in image data, can be extracted. Additionally, residual connections are incorporated into the model. The vanishing gradient problem can be effectively mitigated as the information is allowed to flow directly from shallower to deeper layers by these connections. More effective feature representations can be learned by the network during training through facilitating the backpropagation of gradients. In terms of performance, PPLCNet has distinct advantages. High computational efficiency is exhibited, and the data can be processed rapidly.

2.3. Convolutional block attention module

CBAM (Convolutional Block Attention Module) is an attention model in the field of deep learning.It is designed to adaptively adjust the feature weights.It has the importance of input features learned.So that the key features can be enhanced and the irrelevant ones can be weakened.Thus the overall performance of the neural network can be improved. It can be seamlessly embedded into various convolutional neural network architectures and is utilized to handle data with spatial structures such as images.

Its structure is composed of a channel attention module and a spatial attention module. The channel attention module is focused on the channel dimension of the feature map. Descriptors are obtained through global average pooling and max pooling being carried out, and then they are processed by a shared multilayer perceptron. After that, the channel attention weights are generated through the Sigmoid activation function being applied, and the channels are weighted accordingly. The spatial attention module is centered on the spatial dimension.A spatial attention map is generated through channel concatenation and convolution operations being performed.Then the weights are obtained through the Sigmoid activation function being activated.Finally, the spatial positions are weighted.

Based on the attention mechanism, during the forward propagation, CBAM is processed by these two modules in sequence. The key information in the channel and spatial dimensions can be focused on. Compared with traditional CNNs, a stronger feature representation ability is possessed by CBAM. It is widely applied in computer vision tasks such as image classification, object detection, and semantic segmentation.

2.4. Pedestrian attribute (PA) 100K dataset

PA100K is a dataset in the field of pedestrian attribute recognition. It is large-scale and highly influential. It occupies an important position in the relevant research and application process. It is carefully collected from outdoor surveillance cameras and contains as many as 100,000 pedestrian images. In the current scope of pedestrian attribute recognition data resources, its scale is outstanding. The internal structure of the dataset is meticulously planned. Among them, 80,000 images are designated for training models. The models build effective recognition capabilities through learning from a large amount of data. 10,000 images are used as validation images. These validation images are used to finely adjust and optimize the model parameters during the training process. Another 10,000 images serve as test images. These test images are used to accurately evaluate the performance and accuracy of the model in practical application scenarios.

It is particularly noteworthy that each image is accompanied by 26 detailed and highly representative attribute annotation information. In terms of clothing and accessories, it distinguishes whether a hat or glasses are worn in detail. The types of tops are subdivided into short-sleeved, long-sleeved, striped, logo-printed or patterned, color-blocked, plaid shirts and other styles. The lower garments also cover categories such as striped, printed, long coats, long pants, short pants, skirts or dresses. At the same time, it clearly annotates the presence or absence of shoes and the specific types of bags. Such as handbags, shoulder bags, backpacks, and even whether there are handheld items in front of the body. From the perspective of character characteristics, the age is divided into three segments: over 60 years old, 18 - 60 years old, and under 18 years old. The gender information is clearly defined, and the orientation of the human body is precisely defined as facing forward, sideways or backward. These annotation dimensions are rich and comprehensive.They provide an extremely detailed and accurate data foundation for pedestrian attribute recognition research.They strongly promote the in-depth development and innovative application of technologies in this field.

3. Experimental Design

3.1. Model training

In the research work of this chapter, we will focus on the ppclenet network architecture for in-depth exploration and experimentation. Among them, the depthwise separable convolution module plays an extremely crucial role in the ppclenet network. Through its unique structural design, it effectively reduces the computational load of the network and improves the operational efficiency. To further optimize the performance of this network, we plan to introduce the CBAM attention mechanism into the depthwise separable convolution module respectively. The CBAM attention model can perform adaptive weighting processing on the feature information in the channel dimension. The CBAM attention model can also perform adaptive weighting processing on the feature information in the spatial dimension. It enables the network to more accurately focus on the regions and features in the image that are of critical significance for the task. It significantly enhances the network's ability to extract and represent features.

We employ the PA100K dataset to conduct this experiment. The dataset contains 100,000 pedestrian images collected from outdoor surveillance cameras . And it is one of the relatively large and highly valuable datasets in the field of pedestrian attribute recognition. It is randomly divided into 80,000 training images, 10,000 validation images and 10,000 test images. And each image is annotated with 26 commonly used attributes covering multiple dimensions from clothing to posture. During the specific experimental operation process, we will systematically add the CBAM attention model in different positions of the depthwise separable convolution module in turn. When added in the pre-stage of depthwise separable convolution, observe its effect on attention guidance in the initial feature extraction process. When added in the middle position, explore its influence on the feature transformation and transfer process. When added in the post-stage, analyze its role in the final feature integration and output stage. The specific experiments are shown in Table 1. A meticulous and comprehensive experimental setup is carried out. The PA100K dataset is utilized. A rigorous evaluation and analysis of the models trained under each addition position is conducted in multiple evaluation metrics. The optimal position of the CBAM attention model in the depthwise separable convolution module of the ppclenet network is determined. The improvement of network performance is maximized. A solid and efficient network model foundation for subsequent applications and research in related fields is laid. The development and progress of the entire technical system in related tasks such as image recognition and object detection is promoted.

Table 1. Experimental Setup

Experimental	CBAM Positions
Experiment One	The Second Module
Experiment Two	The Third Module
Experiment Three	The Fourth Module
Experiment Four	The Fifth Module
Experiment Five	The Sixth Module

3.2. Evaluation results

In this experiment, the position of the CBAM attention model within the depthwise separable convolution module

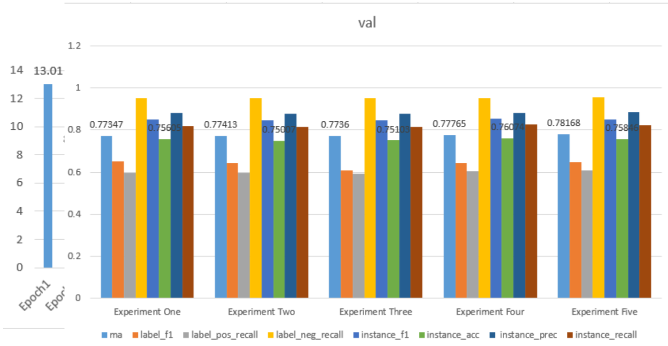


Fig. 5 Experiment Five

Fig. 6 val

was explored intensively. Then, the position was adjusted. Based on this, the training work was carried out. Also based on this, the validation work was carried out. The epoch was set as 40, and the learning rate was precisely set at 0.01. Meanwhile, the momentum parameter was taken as 0.9 to ensure the stability and effectiveness of the training process. During the training process, the changes in loss values under different position settings were closely monitored, and corresponding charts were drawn.

A detailed analysis of the charts is conducted. It enables the clear discernment of performance differences of the CBAM attention model at various positions within the depthwise separable convolution module..Fig. 1, Fig. 2, Fig. 3, Fig. 4 and Fig. 5 respectively display the scenarios where the loss declines as epochs increase for different positions of CBAM in the five experiments.

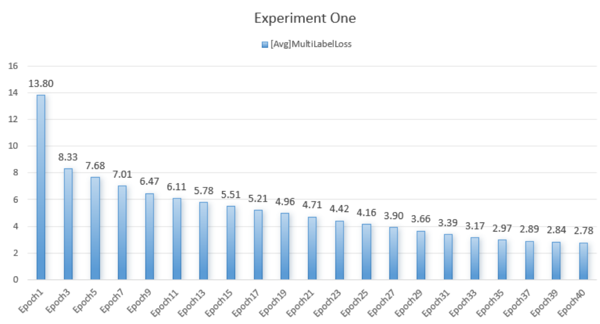


Fig. 1 Experiment One

It can be observed from the aforementioned five figures that after 35 epochs, the rate at which the loss is decreased with the increment of epochs is extremely slow. In Experiment One, a loss value of 2.78 is achieved. In Experiment Two, a reduction of the loss to 2.8 is attained. In Experiment Three, the loss is brought down to 2.77. In

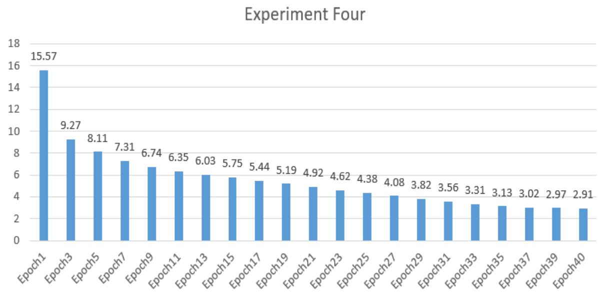


Fig. 4 Experiment Four

Experiment Four, the loss is lowered to 2.91. And in Experiment Five, the loss is diminished to 2.94.

The following Fig. 6 presents the performance outcomes of each group of experiments during the evaluation phase. Our primary focus lies in the mean average (ma) and instance accuracy (instance_acc). Among the five groups of experiments, in terms of ma, Experiment Five exhibits the most outstanding performance with a value of 0.78168. Experiment Four follows closely behind, with a ma value of 0.77765. Regarding instance_acc, Experiment Four shows the most remarkable result, reaching a value of 0.76074. Experiment Five is the runner-up, with an instance_acc value of 0.75846. This indicates that different experimental setups have varying degrees of influence on these two key evaluation metrics. And a comprehensive analysis of such differences is crucial for a more in-depth understanding and optimization of the experimental model and its associated techniques.

4. Conclusion

In this article, the depthwise separable convolutional neural network, the CBAM attention model and the dataset pa100k were first introduced. Then, the attention model was added into different modules of pplcnet for training and evaluation. Through the comparative analysis of evaluation model data, it was discovered that the performance in Experiment Four and Experiment Five was

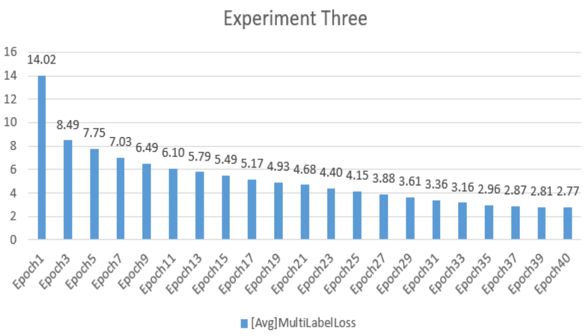


Fig. 3 Experiment Three

better. Namely, when the CBAM attention was placed in the last two modules, a higher model accuracy was obtained. The depthwise separable convolutional neural network served as the base framework. With its unique structural features, it can maintain certain feature extraction ability while reducing computational complexity and has drawn much attention. The CBAM attention model, owing to its dual attention mechanism in channel and spatial dimensions, can enhance the network's ability to capture key information. The dataset pa100k was used to supply data for model training and evaluation.Subsequently, the attention model was inserted into pplcnet modules for training and assessment, involving parameter adjustment and model optimization. A strict experimental process and a comprehensive evaluation were carried out. From the data comparison of evaluation models, it was clearly observed. Experiment Four and Five exhibited significant indicator advantages. Specifically, with the CBAM attention in the last two modules, a higher model precision could be attained. This finding is anticipated to offer valuable reference and guidance for related model architecture design and optimization and promote the further development of research in this area.

Acknowledgements

The research is supported by the major special project of carbon neutralization technology in Tianjin, China in 2024 (24ZXTKSN00070).

References

1. Wu Peishu, Zidong Wang, Han Li et al., KD-PAR: A Knowledge Distillation-Based Pedestrian Attribute Recognition Model with Multi-Label Mixed Feature Learning Network. *Expert Systems with Applications* ,2023.121305.
2. Ji Zhong, Zhenfei Hu, Erlu He, et al., Pedestrian Attribute Recognition Based on Multiple Time Steps Attention. *Pattern Recognition Letters* 138.2020.07.018.

Authors Introduction

Ms. Peng Wang

She is a postgraduate tutor of Tianjin University of Science and Technology. In 2014, she received a doctorate from North China Electric Power University. The research direction is the functional safety assessment of safety instrumented systems.

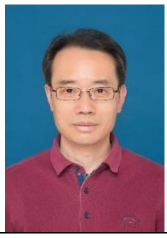
Mr. Shengfeng Wang

In 2023, he received his Bachelor of Engineering degree from the School of Electronic Information and Automation, Tianjin University of Science and Technology, China.He is pursuing a master's degree in engineering from Tianjin University of Science and Technology.

Ms.Yuting Zhou

In 2023, she entered Tianjin University of Science and Technology.She is pursuing a Bachelor of Engineering degree in the School of Electronic Information and Automation.

Mr. Qikun Wang



In 1996, he received his Bachelor of Engineering degree from the School of Electronic Information and Automation, Tianjin University of Science and Technology, China. He is a senior